Analysis of the Binary IV Model

or

The case of the missing 9 dimensions

Thomas Richardson

Department of Statistics

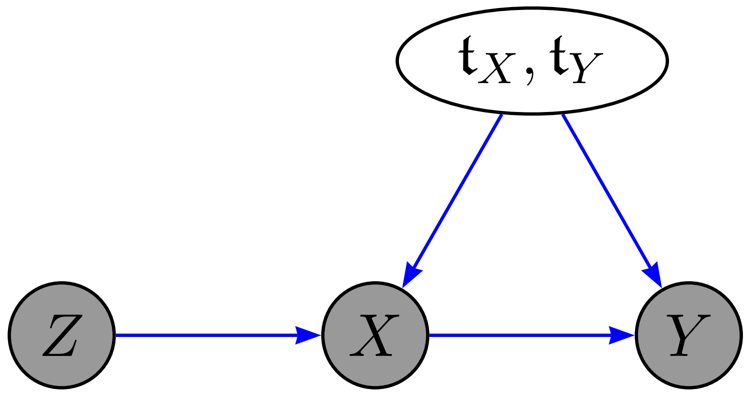
University of Washington

This is joint work with James Robins (Harvard).

Outline

1. The binary IV potential outcomes model
   - *'partial' identification*
2. Even simpler: $X \to Y$
3. Analysis of the binary IV model
   - *pictures*
   - *separating the identified from the unidentified*
4. ?Implications for Bayesian Inference?

Binary IV model



Z is assigned treatment; X is received treatment; Y is the final outcome

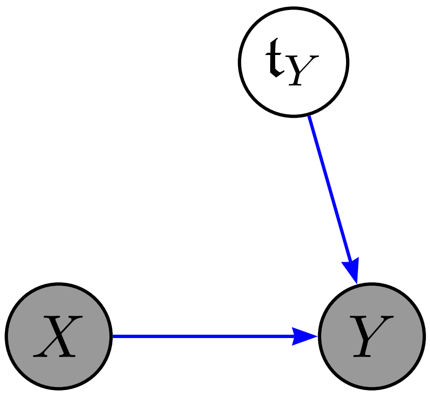
$\mathfrak{t}_X \in \{$Always Taker, Never Taker, Complier, Defier $\} \equiv \{$AT, NT, CO, DE$\}$

$\mathfrak{t}_Y \in \{$*Always Recover, Never Recover, Helped, Hurt*$\} \equiv \{$*AR, NR, HE, HU*$\}$

$p(\mathfrak{t}_X, \mathfrak{t}_Y)$ lives in 15-dim. simplex.

$p(x, y|z)$ lives in a 6-dim. space (product of two 3-dim. simplices).

Even Simpler model: $X \to Y$

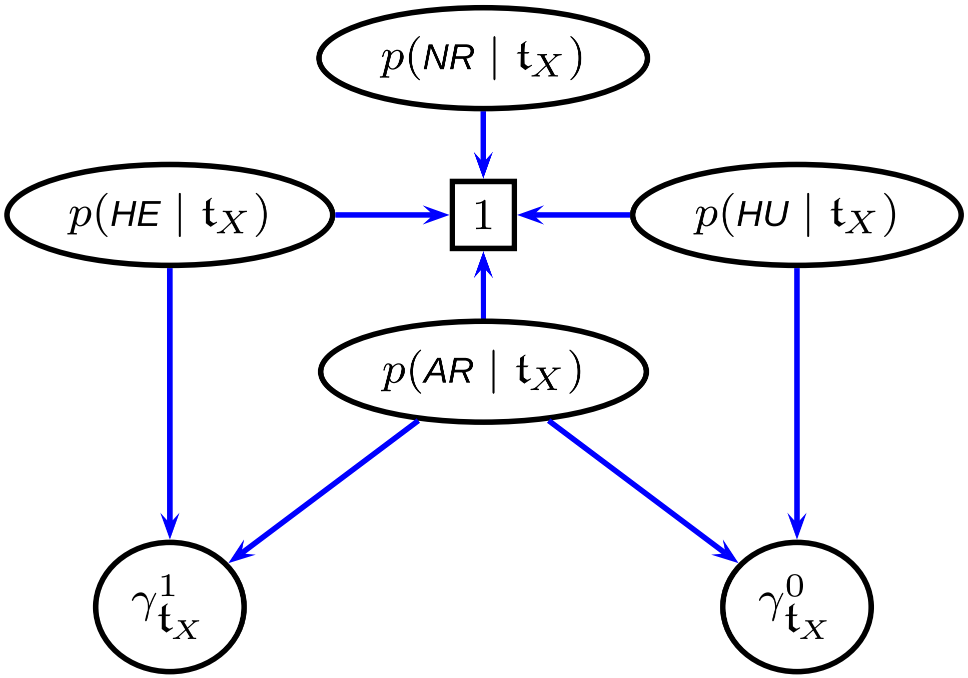


$p(y|x)$ is 2-dimensional

$p(\mathfrak{t}_Y)$ lives in 3-dim. simplex

Set of possible distributions $p(\mathfrak{t}_Y)$ compatible with $p(y|x)$ is of dimension 1.

Cannot determine 'causes of effects'



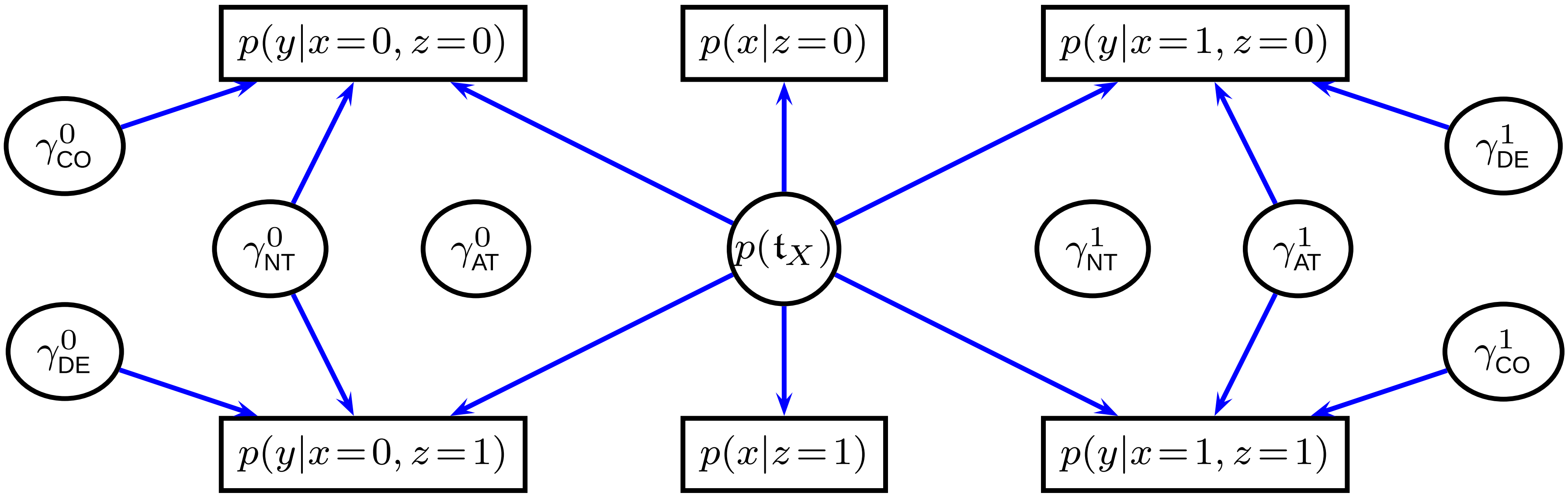
Here $\gamma^0_{\mathfrak{t}_X} \equiv p(y = 1 \mid \mathfrak{t}_X, do(X = 0)),$

$\gamma^1_{\mathfrak{t}_X} \equiv p(y = 1 \mid \mathfrak{t}_X, do(X = 1)).$

This accounts for 4 missing dimensions:

one dimension for ‘causes of effects’ for each compliance ‘type’

This leaves just 5 to go!



Here γ^1_{NT} and γ^0_{AT} are completely unrestricted. (2 dimensions)

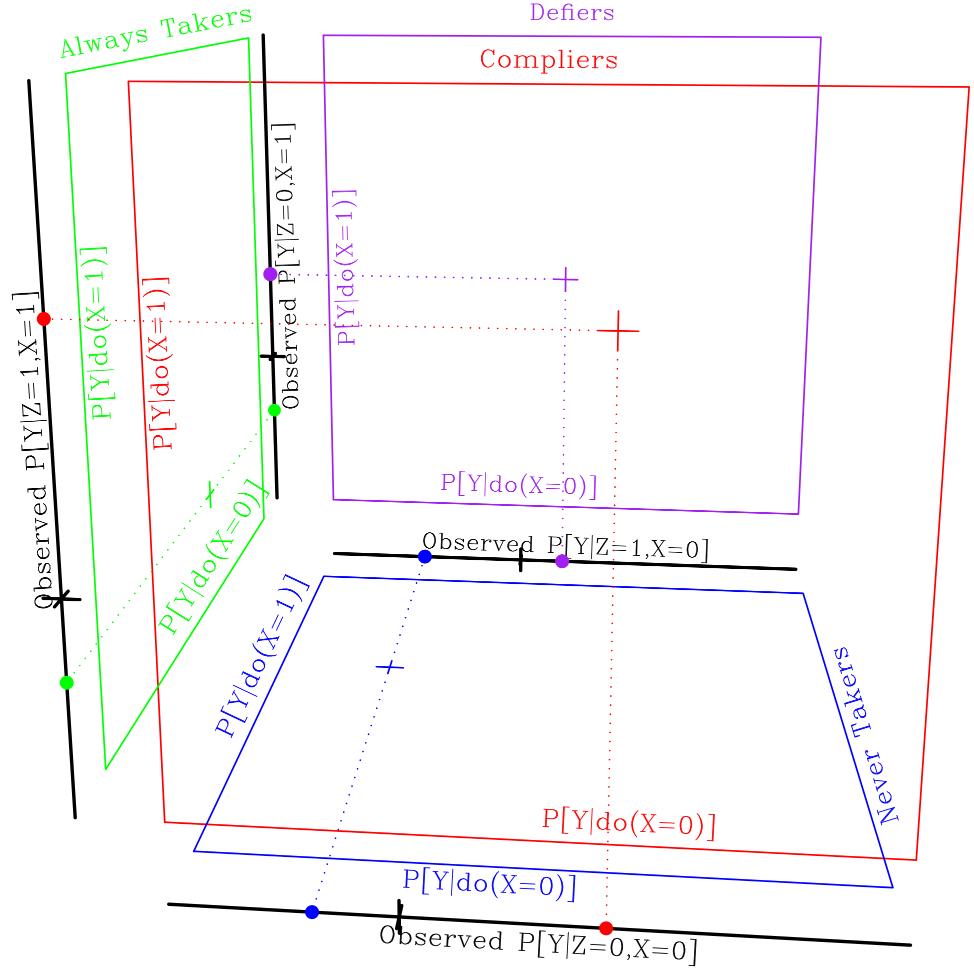
One dimension is associate with each of:

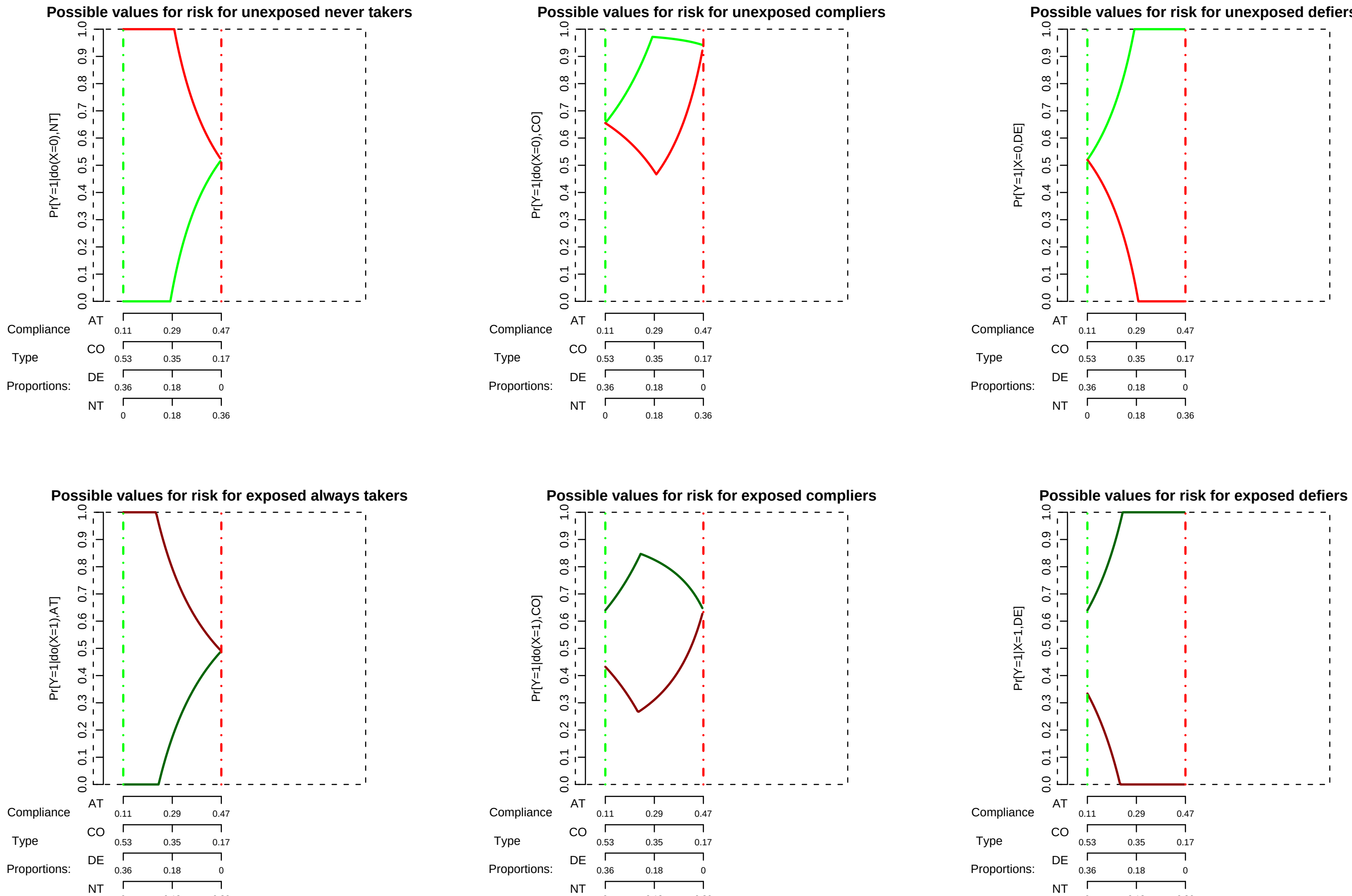
$(\gamma^0_{\text{CO}}, \gamma^0_{\text{DE}}, \gamma^0_{\text{NT}})$

$(\gamma^1_{\text{CO}}, \gamma^1_{\text{DE}}, \gamma^1_{\text{AT}})$

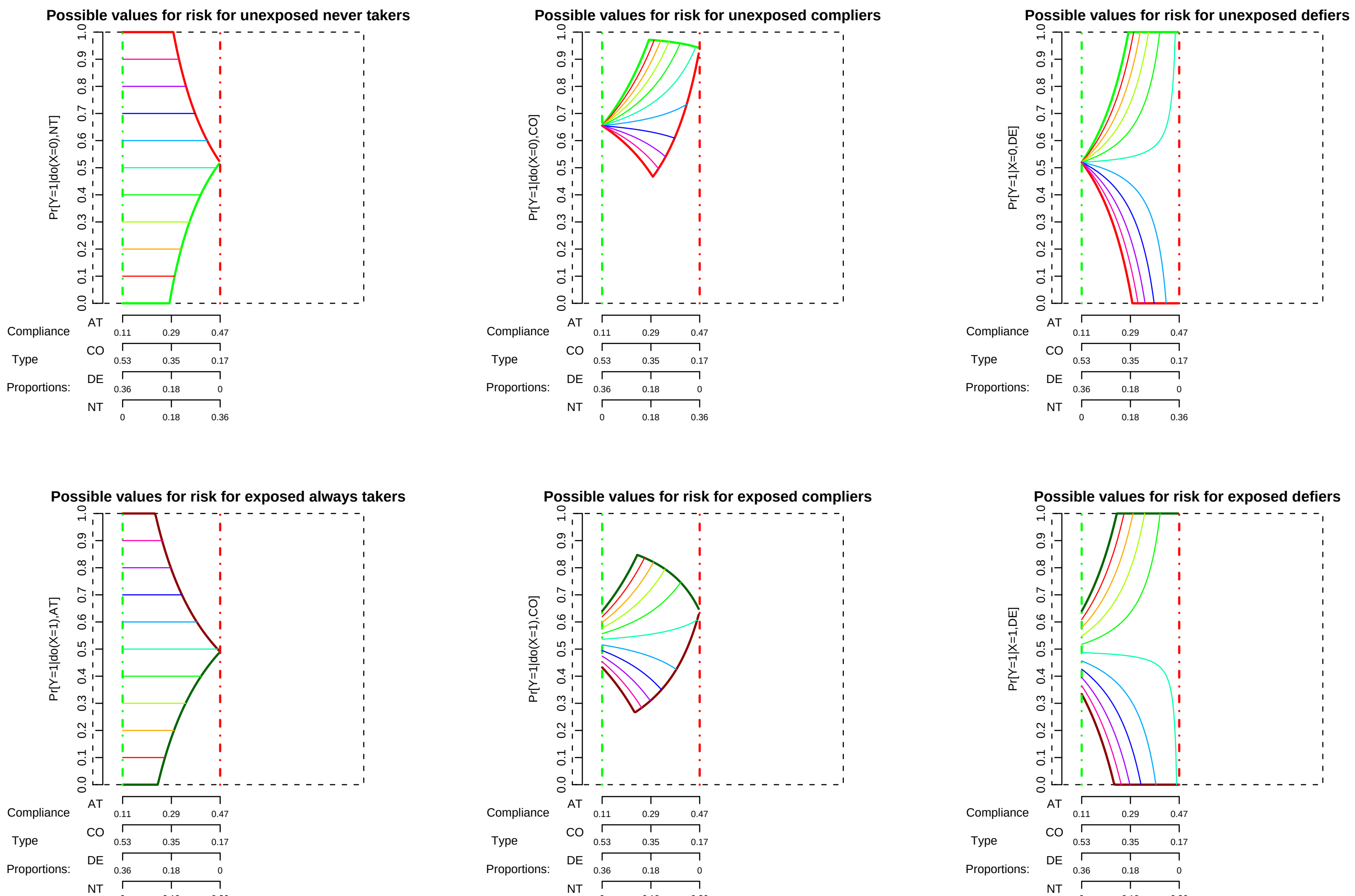
$p(\mathfrak{t}_X) \equiv$ distribution over compliance types

This accounts for the remaining 5 dimensions.

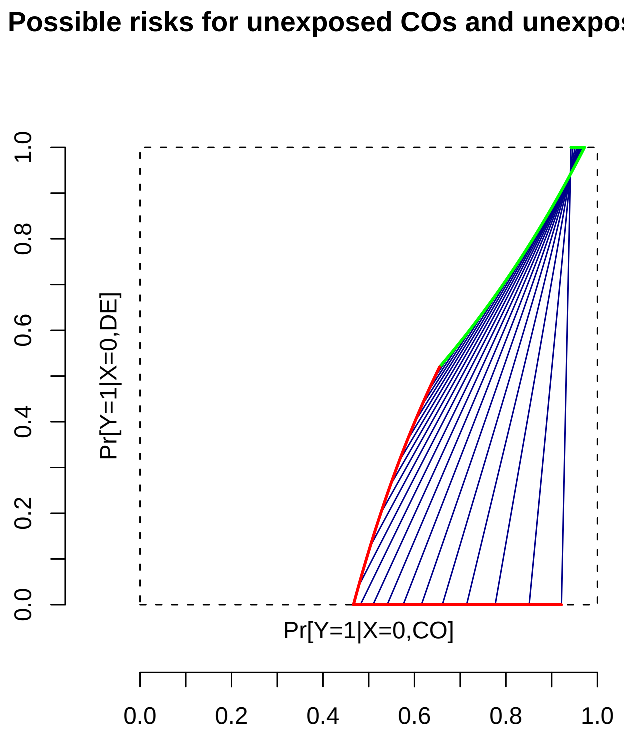

Always Takers
Defiers
Compliers
Observed P[Y|Z=1,X=1]
P[Y|do(X=1)]
P[Y|do(X=1)]
Observed P[Y|Z=0,X=1]
P[Y|do(X=1)]
P[Y|do(X=0)]
P[Y|do(X=0)]
Observed P[Y|Z=1,X=0]
P[Y|do(X=1)]
Never Takers
P[Y|do(X=0)]
P[Y|do(X=0)]
Observed P[Y|Z=0,X=0]

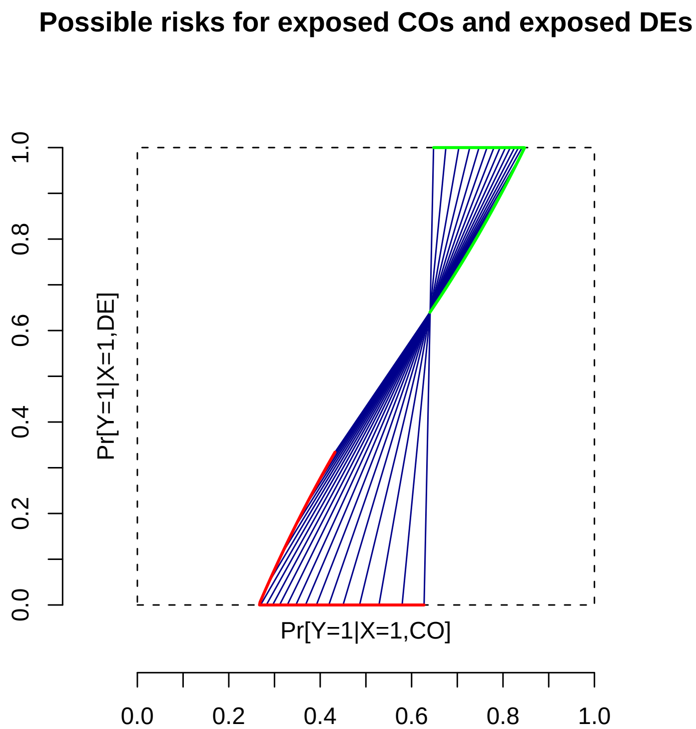


Depiction of the set of values for π_X vs. $\langle \gamma^0_{\mathrm{CO}}, \gamma^0_{\mathrm{DE}}, \gamma^0_{\mathrm{NT}} \rangle$ (upper row), and π_X vs. $\langle \gamma^1_{\mathrm{CO}}, \gamma^1_{\mathrm{DE}}, \gamma^1_{\mathrm{AT}} \rangle$.

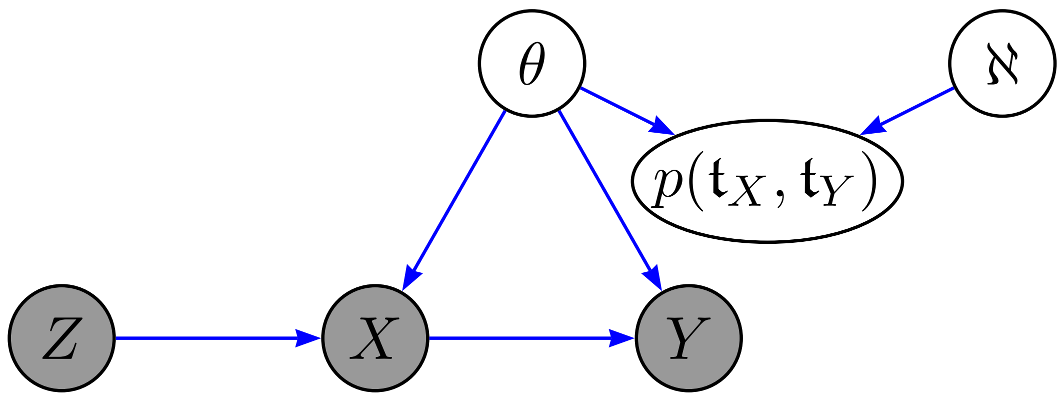


Depiction of the set of values for π_X vs. $\langle \gamma^{0}_{\text{CO}}, \gamma^{0}_{\text{DE}}, \gamma^{0}_{\text{NT}} \rangle$ (upper row), and π_X vs. $\langle \gamma^{1}_{\text{CO}}, \gamma^{1}_{\text{DE}}, \gamma^{1}_{\text{AT}} \rangle$.


Possible risks for unexposed COs and unexposed DEs
Pr[Y=1|X=0,DE]
Pr[Y=1|X=0,CO]
0.0
0.2
0.4
0.6
0.8
1.0


Possible risks for exposed COs and exposed DEs
Pr[Y=1|X=1,DE]
Pr[Y=1|X=1,CO]
0.0
0.2
0.4
0.6
0.8
1.0

Separating identified from non-identified

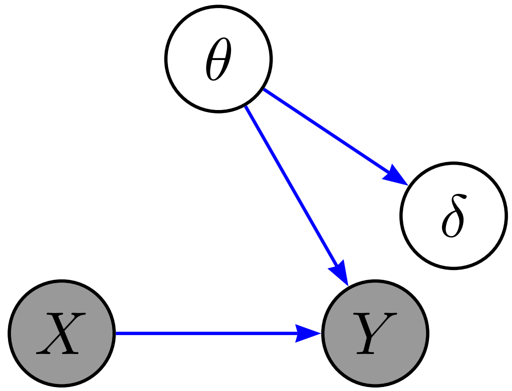


θ is a 6 dim. parameter, (completely!) identifiable from $p(x, y|z)$.

$\aleph$ is a 9 dim. parameter, (completely!) non-identifiable.

$p(\mathfrak{t}_X, \mathfrak{t}_Y) = f(\theta, \aleph)$.

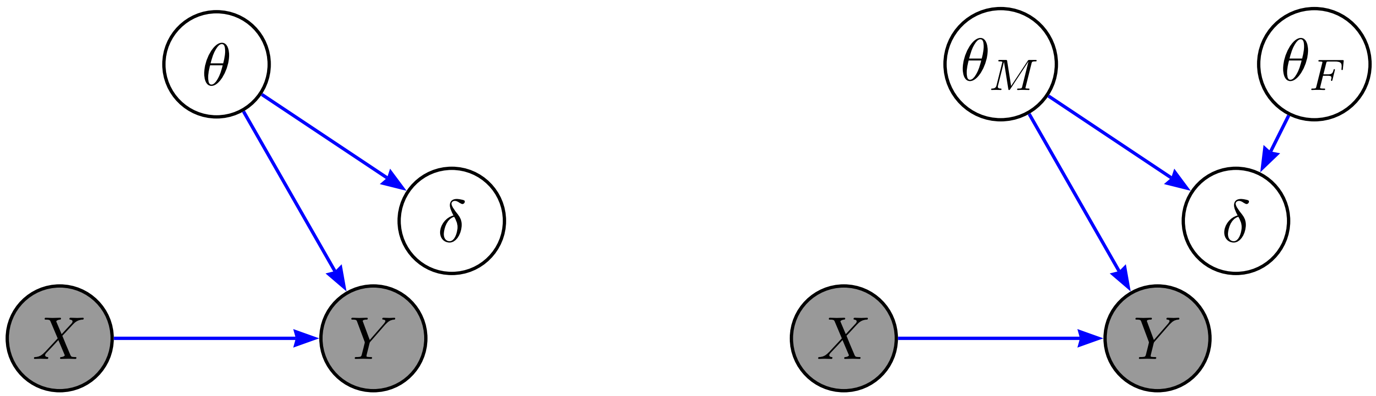
Partial identification: an analogy



- Randomized study to find average effect (δ) of X on Y in a target population containing (a known proportion of) males and females.
- Researcher reports priors on model params (θ), and posterior for δ
  - Aware of partial identification

    but posterior is clearly different from prior $\Rightarrow$ 'informed' by data

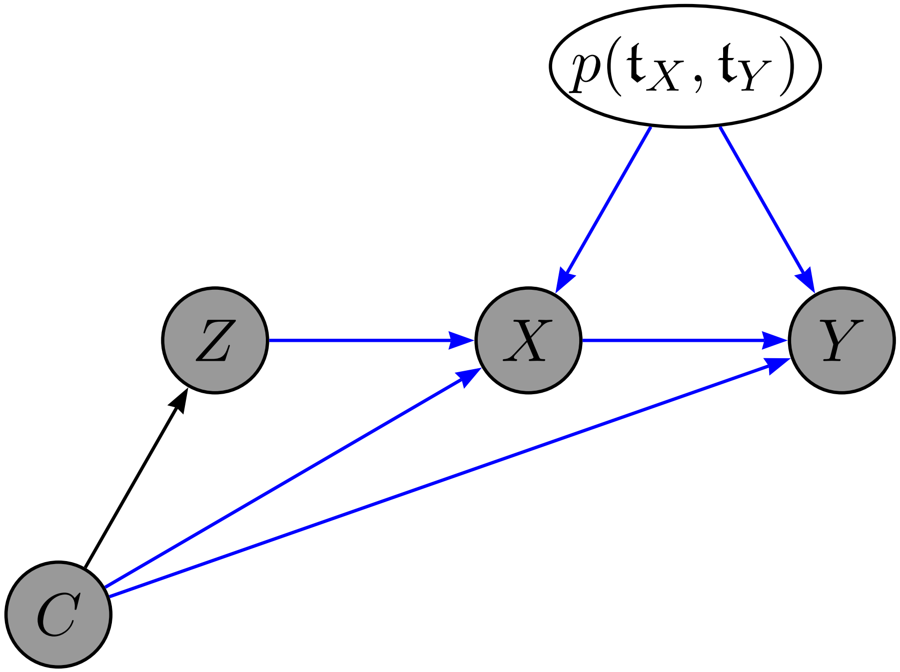
Partial identification: an analogy



- Randomized study of the effect of a drug X on outcome Y in a target population of males and females.
- Later you learn that sample contained no females!

partial identification + proper priors = opaque inference

Future work: incorporating covariates



An identified parameterization makes it possible to parameterize the IV model $p(x, y|z, c)$ conditional on baseline covariates (C).

Parameters For Plots

$$
\begin{aligned}
P(X = 1|Z = 0) &= 0.47 \\
P(X = 1|Z = 1) &= 0.64
\end{aligned}
$$

$$
\begin{aligned}
P(Y = 1|X = 0, Z = 0) &= 0.655 \\
P(Y = 1|X = 0, Z = 1) &= 0.52 \\
P(Y = 1|X = 1, Z = 0) &= 0.49 \\
P(Y = 1|X = 1, Z = 1) &= 0.53
\end{aligned}
$$